
Solucionario de

ejercicios de autoevaluación

Ejercicios de autoevaluación

Unidad de Aprendizaje 1

1. Señala si las afirmaciones siguientes son verdaderas o falsas:

a. La minería de datos es un tipo de analítica de datos predictiva.

- Verdadero
- Falso

b. La minería de datos se ayuda de técnicas como la inteligencia artificial.

- Verdadero
- Falso

c. La minería de datos se basa en utilizar patrones creados por otras técnicas para extraer conocimiento.

- Verdadero
- Falso

d. Los datos no estructurados son la fuente principal de análisis de la minería de datos.

- Verdadero
- Falso

2. Señala cuáles de las siguientes afirmaciones son correctas.

a. El proceso KDD es una de las fases de extracción de conocimiento DM.

b. La metodología SEMMA fue creada por una empresa prácticamente como complemento a su *software* de análisis de datos.

c. La metodología *CRISP-DM* es una guía de libre distribución basada en KDD.

d. Un proyecto de *Data Mining* debe basarse en alguno de estos tres métodos para implementarse.

3. La metodología *CRISP-DM*...

- a. ... contiene cuatro niveles de abstracción divididos desde **tareas generales a específicas.**
- b. ... contiene seis niveles de abstracción divididas desde tareas generales a específicas.
- c. ... sus niveles de abstracción corresponden con las fases de implantación.
- d. ... cuenta con seis fases de implantación que, a su vez, se dividen en niveles de abstracción.

4. Indica qué afirmaciones sobre la fase de comprensión del negocio son verdaderas y cuáles falsas:

- a. Se deben determinar los objetivos de *DM* antes de los objetivos de negocio.
 - Verdadero
 - **Falso**
- b. Durante la elaboración del plan de proyecto se debe definir el contexto en el que se desarrolla el proyecto.
 - Verdadero
 - **Falso**
- c. Debe ser realizada en su totalidad por la dirección de la organización.
 - Verdadero
 - **Falso**
- d. Se deben traducir los objetivos de negocio en objetivos de *DM*.
 - **Verdadero**
 - Falso

5. Indica cuáles de las siguientes afirmaciones sobre la fase de comprensión de los datos son incorrectas:

- a. Una de las tareas de esta fase es la de creación de nuevos datos.**
- b. En la tarea de recolección de datos se listarán los datos que van a ser utilizados.
- c. No es necesario aportar información técnica de los datos, por carecer de sentido.**
- d. En la fase de exploración de datos, se debe describir la estructura de almacenado de los datos en el *Data Warehouse*.**

6. En la fase de preparación de datos...

- a. ... es posible construir datos a partir de los existentes y unificar elementos de tablas distintas.**
- b. ... es una de las fases más rápidas de llevar a cabo, ya que gran parte del trabajo se ha realizado en la fase de comprensión de datos.
- c. ... es necesario hacer como tarea final la limpieza de datos.
- d. ... no es posible modificar el formato de los datos a menos que la dirección de la empresa dé su permiso expreso.

7. Indica qué afirmaciones sobre la fase de modelado son verdaderas y cuáles falsas:

- a. Por cada objetivo establecido se debe seleccionar una técnica de *DM distinta*.
 - Verdadero
 - **Falso**
- b. Por cada objetivo establecido se debe crear un modelo de análisis distinto.
 - **Verdadero**
 - Falso

- c. En esta fase se debe evaluar el modelo *DM* en cuanto a objetivos de negocios.
 - Verdadero
 - **Falso**

- d. Se debe diseñar un plan de pruebas para los modelos creados.
 - **Verdadero**
 - Falso

8. El modelo de *DM*...

- a. ... es independiente de los objetivos de negocios establecidos.
- b. ... debe dar respuesta a los objetivos de negocios establecidos.**
- c. ... debe contar con un plan de pruebas exclusivo, independiente de otros modelos.
- d. ... debe presentar resultados idénticos independientemente de la técnica de datos utilizada, si cuenta con la calidad suficiente.

9. ¿Qué afirmaciones sobre la fase de evaluación son correctas?

- a. Se debe comprobar si los resultados obtenidos cumplen con los objetivos de negocios.**
- b. Tras la validación del modelo se debe definir una lista de acciones para los próximos pasos.**
- c. Se deben descartar los objetivos planteados en el proyecto y mantener solo el que mejores resultados haya dado su modelo.
- d. En esta fase hay que diseñar un plan de mantenimiento de los modelos.

10. Durante la última fase de implantación *CRISP-DM*...

- a. ... se debe diseñar un plan de mantenimiento del proyecto, detallando los pasos a seguir.**
- b. ... se deben establecer las pautas de mantenimiento para el proyecto por parte de la dirección.
- c. ... es obligatorio realizar una presentación final para el personal implicado en el proyecto.
- d. ... el plan de mantenimiento debe ser automatizado, ya que el personal de mantenimiento no está especializado en dichas tareas.

Ejercicios de autoevaluación

Unidad de Aprendizaje 2

1. Selecciona si las afirmaciones siguientes son verdaderas o falsas sobre los tipos de problemas en *DM*:

a. Existen dos tipos de problemas, supervisados y no supervisados, según el tipo de variable a predecir.

- Verdadero
- **Falso**

b. En los métodos descriptivos hay que tratar de identificar patrones en las variables de análisis.

- **Verdadero**
- Falso

c. En los métodos predictivos se debe seleccionar un conjunto de datos de entrenamiento y otro de prueba.

- **Verdadero**
- Falso

d. El aprendizaje inductivo se basa en la creación de fórmulas matemáticas para explicar un fenómeno.

- Verdadero
- **Falso**

2. Señala cuáles de las siguientes afirmaciones son correctas sobre reglas de asociación:

a. **Se basa en descubrir patrones en las transacciones almacenadas en una base de datos.**

b. El soporte de una regla mide la probabilidad si esta realmente es útil o no.

c. **La confianza mide cómo de fuerte es la relación entre dos ítems.**

d. Las bases de datos transaccionales horizontales distribuyen las transacciones en filas y los ítems en columnas.

3. En la métrica de mejora de confianza o *lift*...

- a. ... si el valor es mayor a 1, los productos se considerarán complementarios.
- b. ... si el valor es igual a 1, la regla cuenta con una fortaleza muy alta.
- c. ... si el valor es menor que 1, la regla es fruto del azar.
- d. ... se mide la fortaleza de la regla analizada.

4. Indica qué afirmaciones son verdaderas y cuáles falsas sobre el método de *Clustering*:

- a. Un centroide es un punto donde se hace máxima la suma de las similitudes o distancias de cada elemento del clúster.
 - Verdadero
 - Falso
- b. La distancia euclídea permite medir la distancia de dos elementos o *clústeres*.
 - Verdadero
 - Falso
- c. Para calcular la distancia entre dos elementos siempre se deben normalizar sus valores previamente.
 - Verdadero
 - Falso
- d. En el método de *Clustering jerárquico* se deben definir previamente cuántos clústeres se desean.
 - Verdadero
 - Falso

5. Indica qué afirmaciones son correctas sobre el método de *Clustering*:

- a. El *Clustering* es un método de clasificación de tipo descriptivo.
- b. La técnica de *Single-Link* considera para la formación de segmentos la máxima distancia entre elementos.

- c. Los métodos de *Clustering* de partición normalmente son iterativos.
- d. Un dendograma representa las jerarquías entre los elementos analizados.

6. El método de regresión lineal...

- a. ... es capaz de predecir tanto variables numéricas como nominales.
- b. ... solo es capaz de predecir variables numéricas.**
- c. ... es un caso singular de predicción, ya que no utiliza datos de entrenamiento para crear el modelo de regresión.
- d. ... es un método de predicción basado en reglas de asociación.

7. Indica qué afirmaciones son verdaderas y cuáles falsas sobre la regresión lineal:

- a. El coeficiente indica el grado de ajuste de un modelo de regresión.
 - Verdadero
 - Falso
- b. No puede predecir una variable que dependa de más de una variable independiente.
 - Verdadero
 - Falso
- c. No es posible predecir variables nominales según el contexto estudiado.
 - Verdadero
 - Falso
- d. Un resultado de este modelo es la creación de una función que representa una recta de tendencia de los valores a predecir.
 - Verdadero
 - Falso

8. El método de clasificación...

- a. ... **utiliza normalización de datos para ciertos algoritmos.**
- b. ... no utiliza normalización de datos, ya que usa variables nominales.
- c. ... clasifica los elementos para predecir siempre variables de tipo binomial (dos valores).
- d. ... utiliza el método de distancias para variables nominales para crear reglas de clasificación.

9. ¿Qué afirmaciones son correctas sobre el método de clasificación?

- a. **Se basa en predecir la etiqueta a la que pertenecen los elementos analizados.**
- b. Al igual que el *Clustering*, se divide en métodos jerárquicos y de partición.
- c. Es necesario utilizar siempre técnicas de normalización para las variables numéricas.
- d. **Mediante la matriz de confusión se mide la precisión de un modelo de clasificación.**

10. ¿Qué técnicas se pueden utilizar durante el preprocesamiento de datos?

- a. Solamente técnicas de selección de variables.
- b. **Técnicas que permiten operaciones con filas y columnas de datos.**
- c. El preprocesamiento de datos no es una fase de *DM*, sino de *CRISP-DM*.
- d. Las opciones a y b son correctas.

Ejercicios de autoevaluación

Unidad de Aprendizaje 3

1. Selecciona si las afirmaciones siguientes son verdaderas o falsas sobre Naive Bayes:

a. Es una técnica que se basa en el cálculo de probabilidad.

- Verdadero
- Falso

b. Está orientada a trabajar con variables numéricas.

- Verdadero
- Falso

c. Las simplificaciones que asume disminuyen la precisión del modelo.

- Verdadero
- Falso

d. Es un algoritmo iterativo.

- Verdadero
- Falso

2. ¿Qué afirmaciones son correctas sobre el método de Naive Bayes?

a. Considera que las variables contribuyen de forma independiente a la clasificación.

b. Es un algoritmo orientado al *Text Mining*.

c. A pesar de trabajar con probabilidades, también es capaz de trabajar con distancias.

d. Trabaja con probabilidades condicionales según el teorema de Bayes.

3. El índice de Gini...

- a. ... es un valor que indica la calidad del árbol de clasificación.
- b. ... es un valor que indica cómo de bien separadas están las categorías de una variable.**
- c. ... no es de utilidad cuando las variables a clasificar son nominales.
- d. ... toma valores desde -1 a 1.

4. Di cuáles de las siguientes afirmaciones son correctas sobre SVM:

- a. Solo trabaja con variables numéricas tanto discretas como continuas.
- b. El valor del margen entre clases marca lo bien separadas que están estas.**
- c. Es posible elevar la dimensión de las observaciones para poder clasificarlas.**
- d. Cuanto mayor número vectores de soporte, mejor será la clasificación.

5. Indica qué afirmaciones son verdaderas y cuáles falsas sobre KNN:

- a. Se basa en el cálculo de distancias para clasificar los datos.
 - Verdadero
 - Falso
- b. El modelo de predicción creado es dinámico y actualizable.
 - Verdadero
 - Falso
- c. No funciona bien con datos incompletos.
 - Verdadero
 - Falso
- d. No es posible predecir variables con más de dos categorías.
 - Verdadero
 - Falso

6. Indica qué afirmaciones son correctas sobre Apriori:

- a. Si un ítem es frecuente, todos los grupos que contengan este también lo serán.
- b. Es un método iterativo.
- c. Utiliza gran cantidad de memoria para el procesamiento interno de datos.
- d. Si un ítem es poco frecuente, todos los grupos que contengan este también lo serán.**

7. El método de *K-Means*...

- a. ... es un método iterativo basado en el cálculo de probabilidad.
- b. ... es un método iterativo de convergencia lenta.
- c. ... el resultado será el mismo con independencia de los centroides escogidos.
- d. ... debe conocer *K* y los centroides.**

8. Indica qué afirmaciones son verdaderas y cuáles falsas sobre el algoritmo EM:

- a. Solamente puede trabajar con distribuciones gaussianas.
 - Verdadero
 - **Falso**
- b. Puede trabajar con variables nominales y numéricas.
 - Verdadero
 - **Falso**
- c. Calcula la probabilidad de pertenencia de los datos a los distintos clústeres.
 - **Verdadero**
 - Falso
- d. Se deben inicializar tanto *K* como la media y la desviación típica.
 - **Verdadero**
 - Falso

9. El método de árboles de clasificación...

- a. ... realiza simplificaciones para el cálculo de probabilidad de las variables.
- b. ... cuenta con variantes que permiten realizar árboles de regresión.**
- c. ... debe trabajar con datos normalizados en caso de ser numéricos.
- d. ... es uno de los métodos más exactos que existen en cuanto a predicción.

10. ¿Qué técnica sería más adecuada para clasificar la población mundial según su calidad de vida?

- a. *K-Means*
- b. *KNN*
- c. *EM*
- d. ***SVM***

Ejercicios de autoevaluación

Unidad de Aprendizaje 4

1. Indica si estas afirmaciones sobre la explotación de un proyecto de *Data Mining* estandarizado son verdaderas o falsas:

a. Mejora la probabilidad de éxito y la facilidad de comprensión del proceso.

- Verdadero
- Falso

b. Proporciona una mejor personalización de las fases a las particularidades de la empresa que lo utilice.

- Verdadero
- Falso

c. Facilita la capacidad de respuesta ante nuevos retos y objetivos.

- Verdadero
- Falso

d. Proporciona los conocimientos necesarios para su implementación sin necesidad de disponer de personal especializado.

- Verdadero
- Falso

2. ¿Quién será el encargado de dirigir y administrar un proyecto de *DM*?

- a. El alto mando.
- b. La gerencia.
- c. La dirección operativa.
- d. La gerencia ejecuta su dirección y la dirección operativa administra los elementos específicos.**

3. Indica cuáles de las siguientes afirmaciones son correctas sobre la primera fase del proceso de elaboración del proyecto de *DM*:

- a. **La dirección de la organización será la encargada de definir los objetivos de negocio.**
- b. **La figura que se encarga del análisis de datos será el encargado de definir los objetivos de *Data Mining*.**
- c. En la tarea de evaluación de la situación se debe definir y desarrollar el contexto de la organización.
- d. Los criterios de éxito siempre deben ir orientados a los objetivos de negocio.

4. Indica qué respuestas son verdaderas y cuáles falsas sobre la segunda fase del proceso de elaboración del proyecto de *DM*:

- a. Esta fase incluye una tarea de verificación de la calidad de los datos.
 - **Verdadero**
 - Falso
- b. En la tarea de selección y limpieza de datos se deben elegir los datos definitivos con los que se trabajará.
 - **Verdadero**
 - Falso
- c. En la tarea de descripción de datos se deben definir los datos que se van a utilizar y describirlos detalladamente.
 - Verdadero
 - **Falso**
- d. Se debe realizar un informe de exploración de datos con análisis estadísticos que aporten información relevante antes del análisis *DM*.
 - **Verdadero**
 - Falso

5. Indica cuál no es una tarea de la fase de preparación de los datos:

- a. Formateo de datos.
- b. Selección y descripción de datos.**
- c. Integración de datos.
- d. Construcción y formateo de datos.

6. Indica qué afirmaciones corresponden a las tareas de preparación de los datos:

- a. En esta fase los datos deben llegar ya limpios para realizar las tareas de selección, construcción y formateo de los mismos.
- b. Se deben construir tantos los datos como las tablas necesarias para su utilización por la herramienta DM.**
- c. Esta fase la debe realizar íntegramente el analista de datos o el equipo técnico designado.**
- d. Se deben definir los *Datasets* que se van a utilizar para cada uno de los objetivos DM.**

7. Indica qué afirmaciones son verdaderas y cuáles falsas sobre la fase de modelado:

- a. Se debe realizar un plan de prueba que explique el conjunto de operaciones que se le van a practicar a los modelos conjuntamente.
 - Verdadero
 - **Falso**
- b. Se deben evaluar los modelos desde un punto de vista del negocio antes de pasar a la siguiente fase.
 - Verdadero
 - **Falso**
- c. Cada objetivo de negocio contará con su propio modelo de análisis.
 - Verdadero
 - **Falso**

d. Tras la construcción del modelo se analizará técnicamente su desempeño.

- Verdadero
- Falso

8. En la fase de modelado:

- a. Se elegirá el *software* con el que se realizarán los modelos *DM*.
- b. Se elegirán las técnicas a utilizar para cumplir cada objetivo de *DM*.**
- c. Se elegirán las técnicas a utilizar para cumplir cada objetivo de negocio.
- d. Se realizará un reporte a la dirección del proyecto sobre los resultados de evaluación de los modelos.

9. ¿Qué acciones no pertenecen a la fase de evaluación?

- a. Evaluación de los modelos en cuanto a los objetivos de negocio.
- b. Evaluación de los modelos en cuanto a los objetivos *DM*.**
- c. Determinación de los pasos siguientes a realizar.
- d. Elegir un modelo definitivo para cada objetivo.

10. Selecciona qué afirmaciones son verdaderas y falsas sobre la sexta fase del proyecto:

- a. Durante la fase de despliegue es tarea del analista de datos o personal técnico diseñar el plan de mantenimiento.
 - Verdadero
 - Falso
- b. También es tarea del analista de datos o personal técnico designado realizar el informe final y la presentación del proyecto.
 - Verdadero
 - Falso

c. El plan de mantenimiento debe estar integrado en el que tenga ya implantado la empresa.

- Verdadero
- **Falso**

d. Un proyecto de *DM* es de índole técnica por lo que su existencia no debe ser publicada al resto de personal no vinculado a él.

- Verdadero
- **Falso**

